

AI破局挑戰：

NetApp 創新資料管理平台

Sapphire Lin 林亞璇

Solution Engineer, NetApp team

Sep. 2025

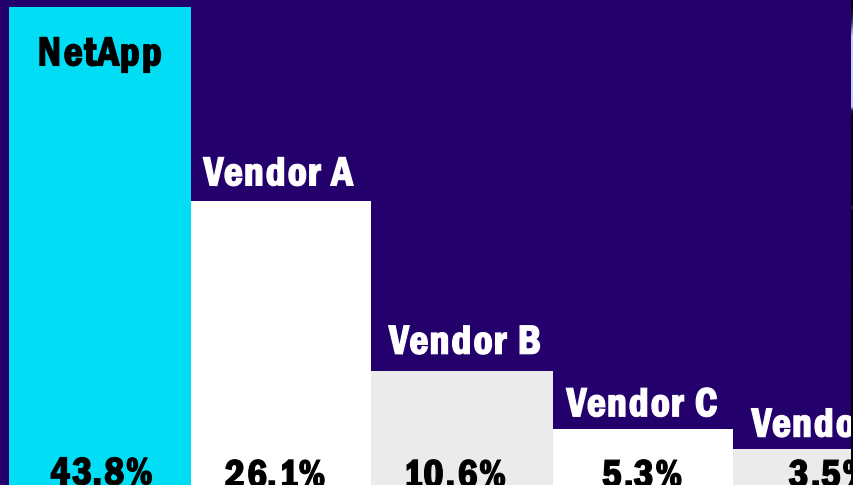


NetApp #1 IN TAIWAN

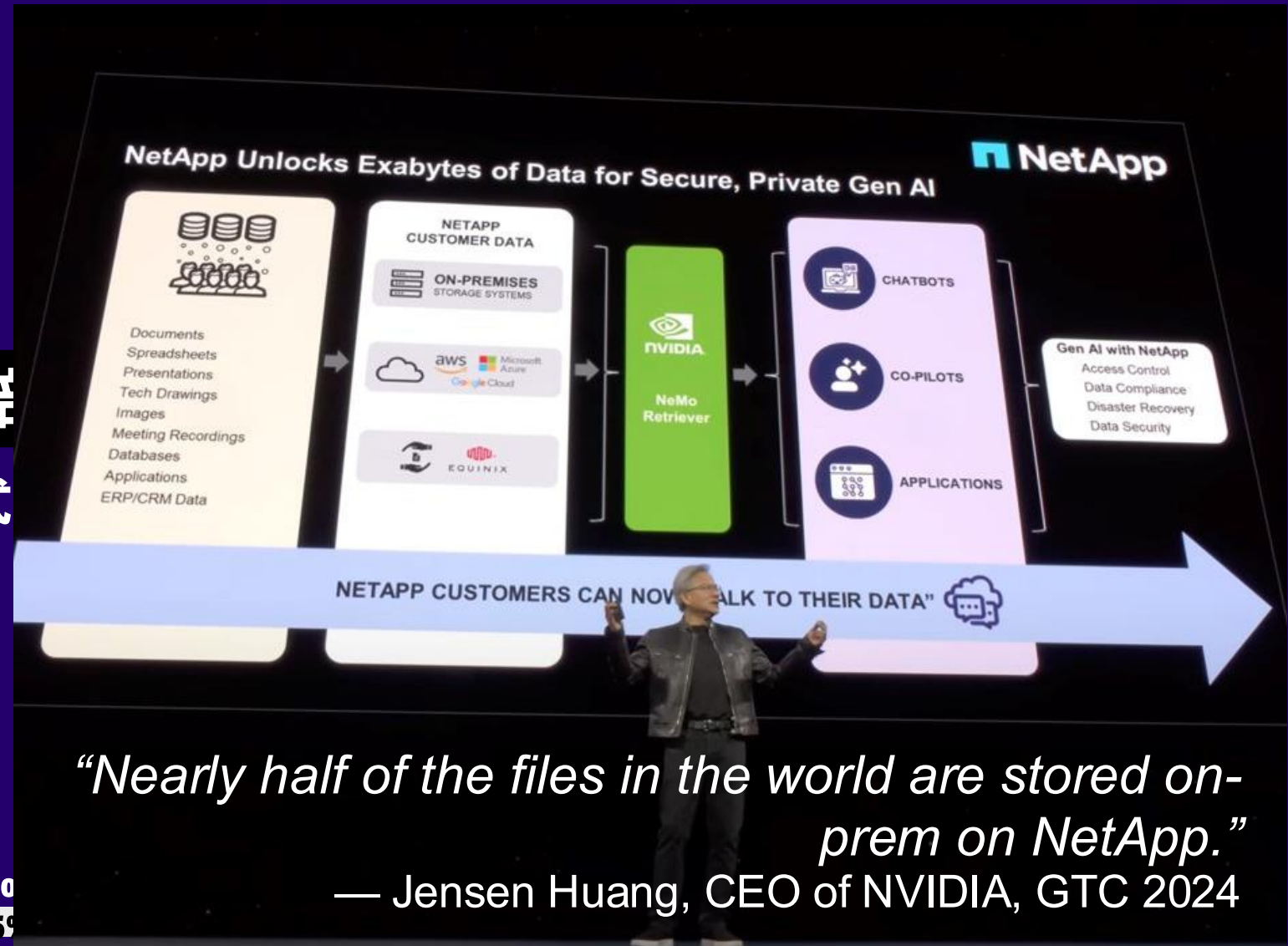
資料來源：IDC 全球企業儲存系統季度追蹤報告, 2025Q1 (2025 年 6 月發布)



企業外
開放式



2025 Q1 廠商市場收入計算



2025 Q1 廠商市場收入計算

跨越多個世代，NetApp 一直引領資料基礎架構創新

資料量、複雜性和成本不斷增加



Adoption



Acceleration

Complexity



Simplification



2002

統一化儲存解決數據孤島

NetApp becomes the first vendor to **unify file & block workloads, and structured & unstructured data**

2012

打通混合雲的最後一哩路

NetApp creates the first **data fabric strategy** that eliminates silos & provides unified control across any environment

2022

三大公雲原生的儲存服務

NetApp becomes the **ONLY** vendor to introduce **cloud ops and data services** as key data infrastructure pillars in addition to being the **only vendor natively embedded in all major clouds**

Today

智慧型基礎架構

NetApp delivers **silos-free infrastructure**, then harnesses **observability and AI** to enable best data management everywhere

BRING AI TO DATA



首先, AI workload 需要 external storage?

資料規模

要讓AI模型準確度越高，訓練用的資料越多越好，在這個前提下，資料量會不斷成長

Local storage空間遲早不夠，所以需要容易擴充的**external storage**。另外，現在的Gen AI LLM模型越來越大，參數越來越多，**GPU memory**過於昂貴空間不夠的狀況下，也需要高效能的**external storage**來幫忙最佳化整體成本。

LLM的典型狀況是，一個具有1 billion (10億) 參數的model，需要GPU memory至少80GB (32-bit full precision)。Meta的LLaMA 2有約70 billions (700億) 參數，估計需要 $80GB \times 70 = 5600GB = 5.6TB$ 的GPU memory，而OpenAI的GPT4則是號稱有兩萬億參數，估計需要 $80GB \times 2000 = 160TB$ 的GPU memory...?

資料分享

多個**GPU node**需要由統一平台提供資料給不同**GPU node**進行運算，避免資料孤島

所以需要分享資料用的**external storage**，解除**local storage**管理資料的限制與麻煩，提升資料科學家的工作效率

資料獨立

AI時代的應用不斷改變，但資料才是企業最寶貴的資產，因此將計算與儲存分開，做到計算不斷迭代，而資料永久持續

這也正是Microsoft, Facebook, AWS, Google, Nvidia等科技巨頭的做法，所以需要獨立運作的**external storage**來達到此目的。

資料安全

如何**最有效率地保護資料**，備份資料、還原資料、移動資料、加密資料、防止勒索病毒、訪問控制、分析使用行為、分類識別隱私等，這些都是非常重要的安全考量，所以需要專業的**external storage**來提供這些企業級服務

AI Workload 流程分析

Gen AI (LLM) Workflow:

Data preparation

數據處理

High performance model training

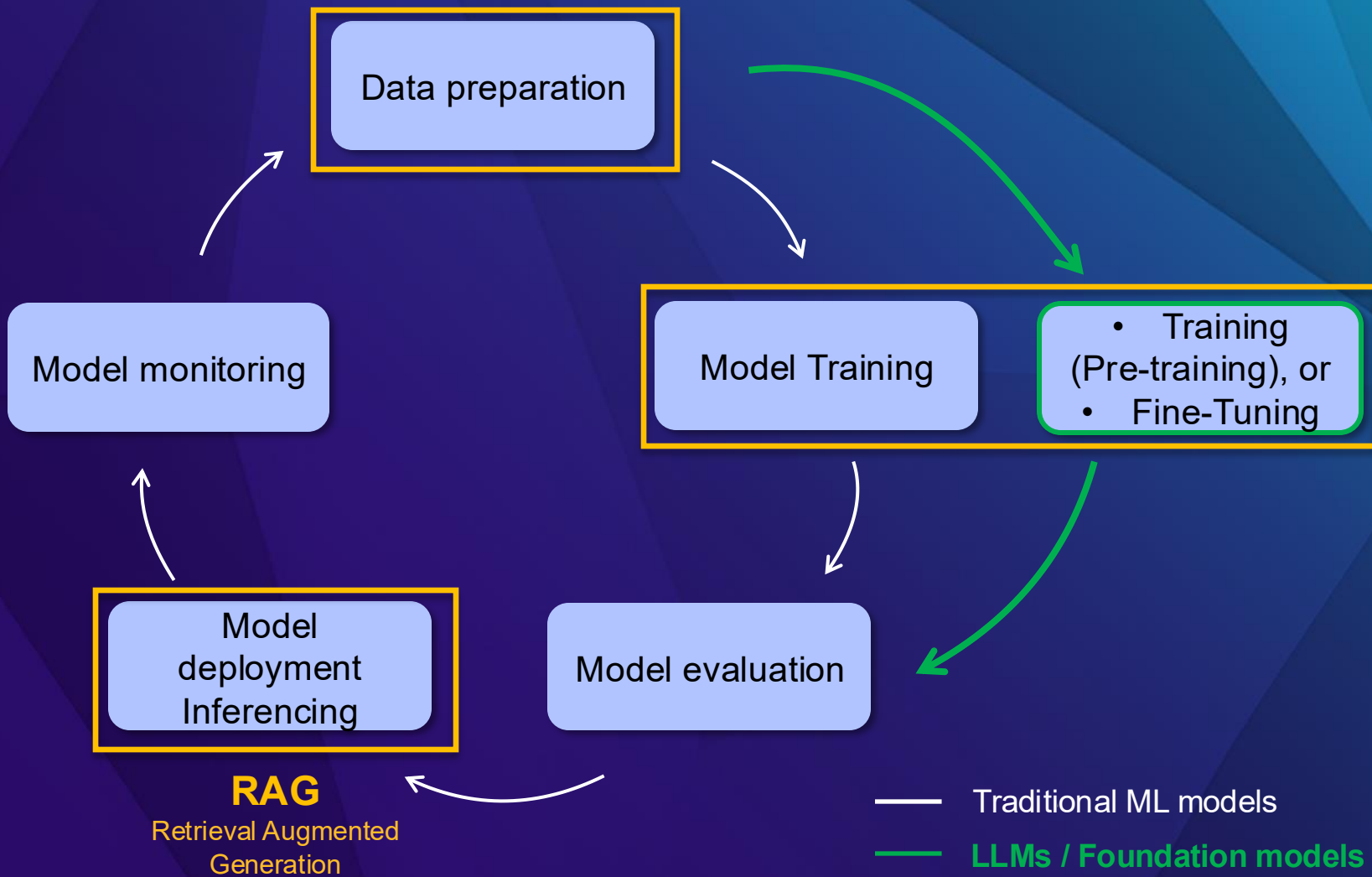
從無到有建立model做訓練

Fine tuning pre-trained model

套用既有model微調

RAG and inferencing

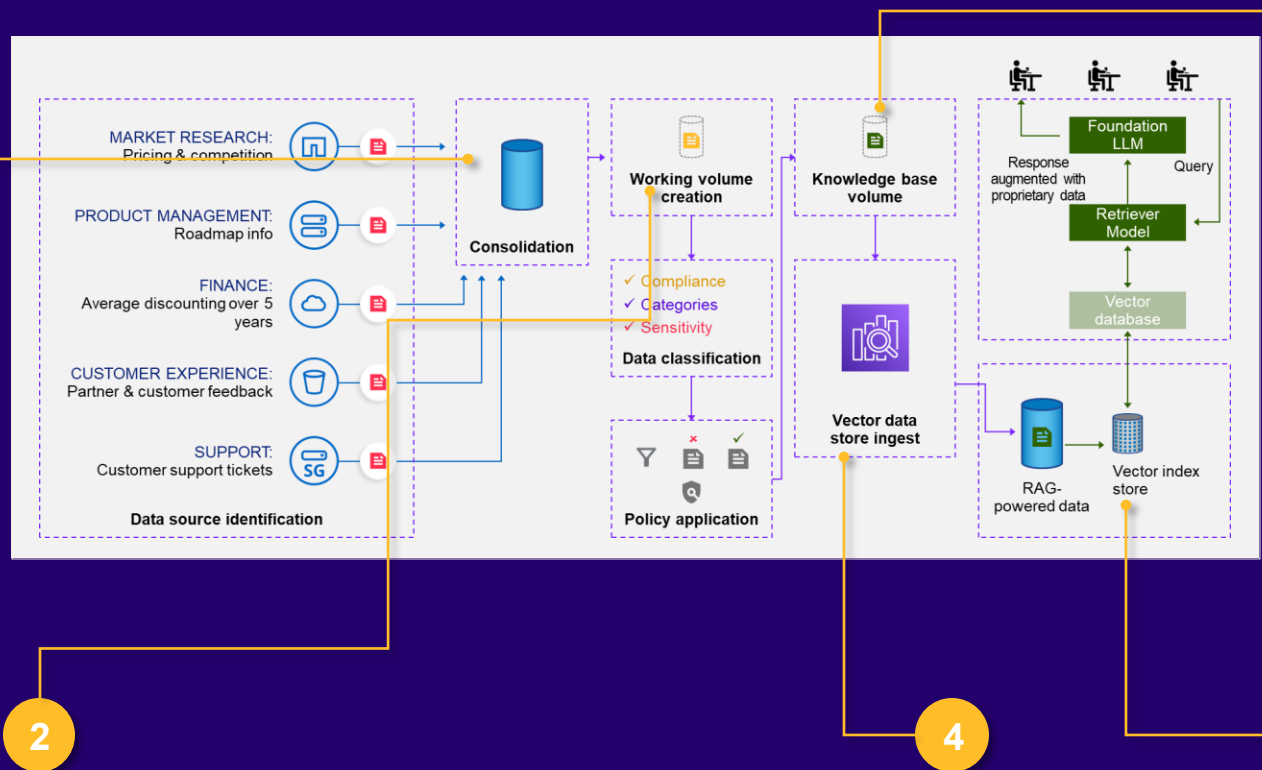
搭配公司knowledge base做推理



使用企業級 RAG 遇到的 5 個典型問題

1

如何彙整來自整個企業的資料——無論它位於何處、也無論其格式？



2

如何確保一致的資料來作業，不讓無關資料污染 GenAI 資料管道，進而降低其相關性？

4

如何以程式設計方式在最具成本效益的地方使用 GPU 資源？

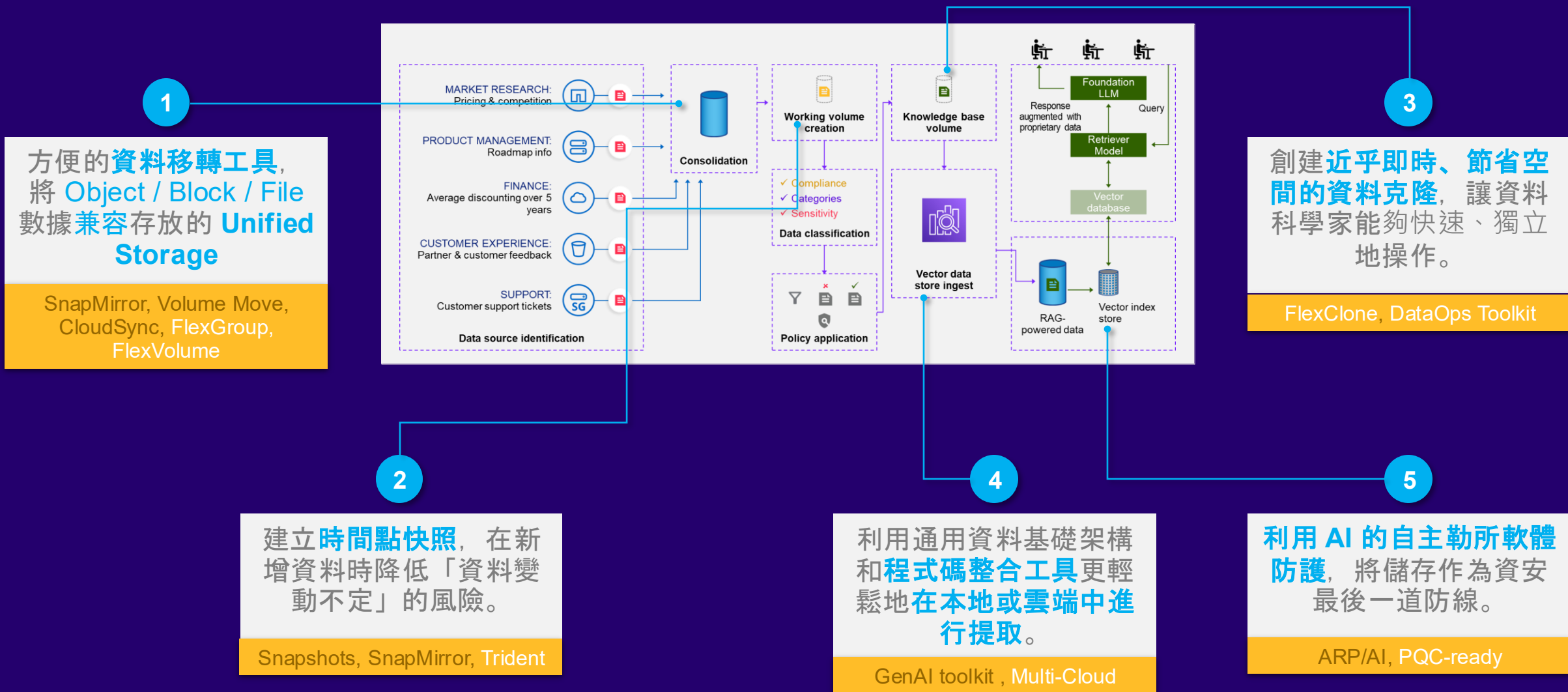
3

資料科學家需要各自獨立的沙盒來進行實驗和測試，我不能讓他們互相干擾，但我又沒預算讓每個人都自己的儲存空間。

5

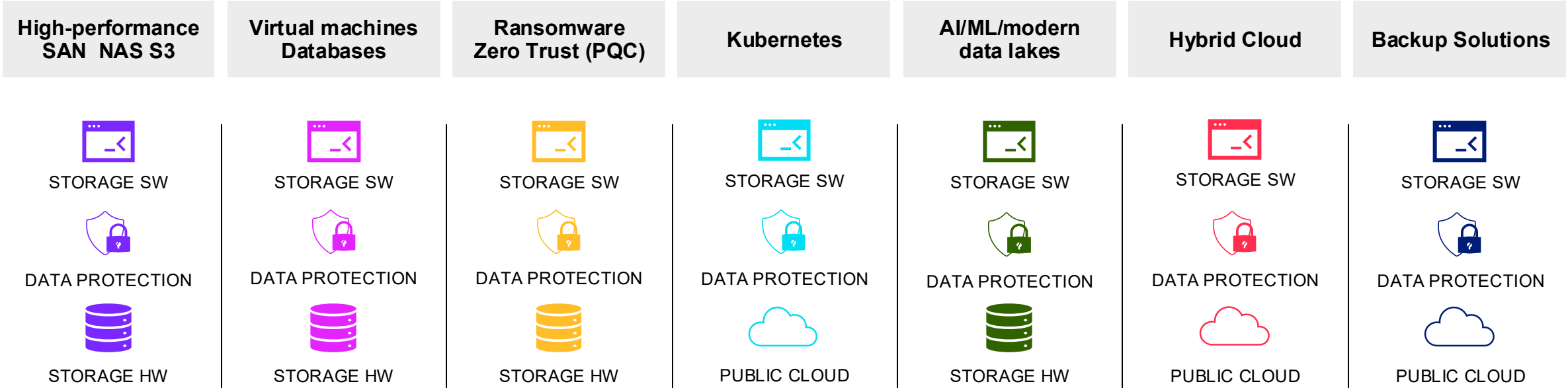
企業級RAG 方案基礎架構需要提供資安韌性（資料保護與可用性），這個應用最終會成為業務關鍵系統，不能讓敏感資料外洩。

使用 NetApp 架構企業級 RAG 的 5 大優勢



複雜性是 IT 生產力的無聲殺手

跨應用程式、資料和雲端的客製化基礎架構孤島效率低、成本高昂，並且會帶來風險



✗ 管理和自動化不一致 ✗ 資料安全不一致 ✗ 資料保護不一致 ✗ 效能表現不一致 ✗ 使用體驗不一致

NetApp 協助您化繁為簡

Any Data, Any workload, Any Storage protocol, Anywhere

vmware®

ORACLE

Microsoft®
SQL Server™

SAP®

High-performance
SAN NAS S3

Virtual machines
Databases

Ransomware
Zero Trust (PQC)

Kubernetes

AI/ML/modern
data lakes

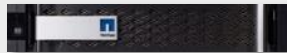
Hybrid Cloud

Backup Solutions

ONTAP

ONLY NETAPP

Comprehensive data management software delivering automation, efficiency, data protection, and security capabilities **for file, block, and object**



aws

Microsoft
Azure

Google Cloud

IBM Cloud

- ✓ 一致性的管理和自動化
- ✓ 一致性的資料安全
- ✓ 一致性的資料保護
- ✓ 一致性的效能
- ✓ 一致性的使用體驗

DATA ACCESS

WITHOUT NETAPP

Filter files by name

Name	Modified	File Size
Examples	25m ago	
netapp_d...	1h ago	
DemoWit...	7m ago	6 KB

Importing of Libraries

```
[ ]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import datetime as dt
```

Documents

liveDemo.csv

```
[ ]: taxiData["fare_amount"].plot(figsize=(27,9), xlim=(dt.date(2017,1,1), dt.date(2017,1,1)))
```

Cleaning of Data

```
[ ]: taxiDataClean= taxiData.copy()
```

NETAPP

Filter files by name

Name	Modified	File Size
Examples	56m ago	
netapp_d...	4m ago	
DemoWit...	2m ago	7.5 KB

Importing of Libraries

```
[ ]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import datetime as dt
from netapp_dataops.traditional import mount_volume
```

DataOps Toolkit | Mounting volume

```
[ ]: mount_volume(volume_name="taxiData", mountpoint="/home/user/taxiData", print_out=True)
```

Reading Data

```
[ ]: taxiData = pd.read_csv("/home/user/taxiData/liveDemo.csv")
taxiData["Unique"] = taxiData.index
taxiData["date"] = pd.to_datetime(taxiData["lpep_pickup_datetime"], format="%Y-%m-%d")
taxiData.set_index("date", inplace=True)
taxiData.sort_values("date", inplace=True)
taxiData.info()
```

Visualization of the Data

Rides per Day

```
[ ]: taxiData.groupby("date").nunique()["Unique"].plot(figsize=(27,9))
```

Costs per Ride

```
[ ]: taxiData["fare_amount"].plot(figsize=(27,9), xlim=(dt.date(2017,1,1), dt.date(2017,1,1)))
```

儲存是最後一道防線

唯一經由美國國安局（NSA）CSfC 認證
存放機密資料的企業級儲存系統



Commercial Solutions for Classified (CSfC)



FIPS 140-2



Department of Defense
Approved Product List
(DoDIN APL)



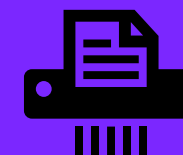
Common Criteria

Policy Engine
黑、白名單



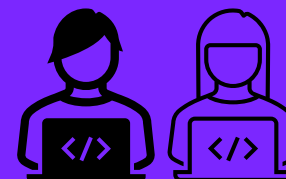
Automatically block known
malicious file types

**Immutable and tamper-
proof Snapshot copies**
防篡改、防刪除快照



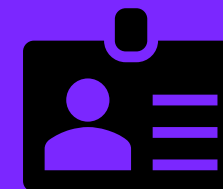
Prevent data destruction with
immutable and indelible
Snapshot copies

Multi-Admin Verify
多管理員驗證



Block rogue admins and
malicious users

Data-in-Transit Encryption
後量子時代加密防護



Secure data access,
end to end, PQC ready

利用 AI 進行自主勒索軟體防護

(ARP/AI)

Next generation ransomware threat detection



世界上第一個針對主要儲存的 AI 驅動型
On-Box 勒索病毒偵測



下一代 AI/ML 從第一天起就確保精確檢測

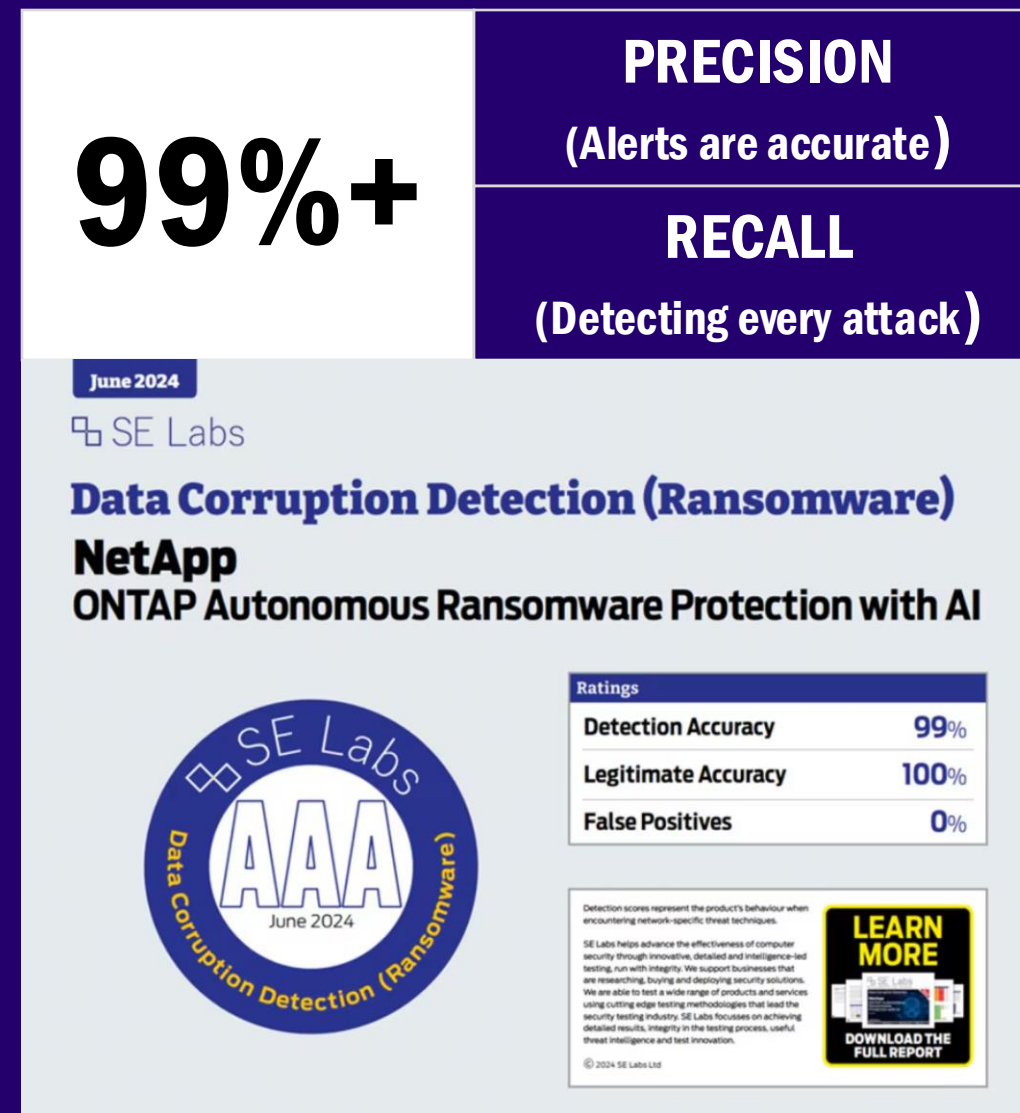


自動更新可提供最佳、高效且準確的一站式
主動偵測



為現有 ARP 客戶輕鬆自動升級

SE LABS rating:
<https://ntap.com/3VIHMF2>



Real-time detection with is essential to effective recovery - ARP/AI

全球首款內建 AI 驅動的即時勒索病毒偵測與回應系統

秒級啟動

Anti-ransomware

STATUS

Enabled in learning mode

Started Time: 16 Sep 2021 6:51:PM

The system has been auto-learning the workload characteristics of t

A learning period of 7 to 30 days is recommended.

You can switch from "learning" to "active" mode at any time.

Switching early might lead to too many false positive results.

Switch to active mode

Monitor Config

Monitor new file types

New file types are file types that were never detected during the learning mode.

MAXIMUM NUMBER OF NEW FILE TYPES THAT'S CONSIDERED NORMAL

20

 in

24

 Hours

事件告警

Event severity settings

Event	Severity
Detected new file extensions	Don't generate ev...
Created a ransomware Snapshot copy	Notice
Detected a ransomware attack	Alert

Cancel

Save

自動觸發快照保護

Overview Snapshot Copies SnapMirror (Local or F

+ Add

Name
Anti_ransomware_backup.2021-11-18_1506
hourly.2021-11-18_1505

條列潛在檔案

Suspected File Types

You can mark the suspected file types as false positive or potential ransomware attack. The system will make necessary adjustments to the way it detects abnormal activity in the future.

Type

lckd

FILES (SHOWING A MAXIMUM OF 20 FILES)

/Top_Dir_1/Sub_Dir_11/test_file_8335.5.png.lckd

/Top_Dir_1/Sub_Dir_11/test_file_27373.5.html.lckd

/Top_Dir_1/Sub_Dir_11/test_file_27373.5.mp4.lckd

/Top_Dir_1/Sub_Dir_11/test_file_27373.5.doc.lckd

/Top_Dir_1/Sub_Dir_11/test_file_27373.5.gif.lckd

/Top_Dir_1/Sub_Dir_11/test_file_103.2.jpg.lckd

/Top_Dir_1/Sub_Dir_11/test_file_103.2.doc.lckd

/Top_Dir_1/Sub_Dir_11/test_file_3040.5.html.lckd

/Top_Dir_1/Sub_Dir_11/test_file_3040.5.pdf.lckd

NetApp PQC Ready



TODAY



DATA-AT-REST

- PQC ready now using AES-XTS-256 for data-at-rest encryption

2025 READY



DATA-IN-TRANSIT

- Scheduled for ONTAP 9.17.1 release
- Lattice-based Key Encryption support for OpenSSL 3.5, ML-KEM, ML-DSA (**FIPS203, FIPS204**)
- Stateless hash-based digital signature support for SLH-DSA (**FIPS205**)



DATA-IN-USE

- In-use encryption includes off-load cards, and in-memory operations
- Working with hardware vendors and partners to develop future solutions

美國海軍 Trident Warrior

AI 與隱私保護演習平台

- 需在不暴露原始影像的情況下，進行即時辨識 VIP 臉部
- 傳統解決方案常需高硬體資源，且在機器學習生命週期中表現不一致，導致部署延遲

NVIDIA + NetApp 解決方案

- NVIDIA TESLA V100 GPU + NetApp AFF GPU Direct Storage
- Stained Glass Transform™ 將影像轉換為僅供 AI 模型解讀的隨機化再表示 (Randomized Re-Representations)
- 部署過程中，資料保留在海軍內部，確保資料主權與安全。

成果與效益

- 超過 99% 高精準臉部辨識
- 部署過程中，資料保留在海軍內部，確保資料主權與安全
- 未來可結合公雲 LLMs 和生成式 AI 模組



國衛院啟用 NetApp AFF A800，為生醫 AI 平台建立快速且流暢的資料傳輸通道

以 A800 為中繼儲存，快速輸送資料至 GPU 伺服器，提升 AI 建模效率

文/ 廠商新聞稿 | 2022-10-14 發表

讚 45

分享



「我認為 NetApp AFF A800 最大優勢，在於擁有優異的資料儲存與分享效率，並擁有高穩定度，且只要我們拋出問題、NetApp都能快速回應。這些特質，足以帶給國衛院充足信心，未來不論承載多麼龐大複雜的基因資料庫，也無所畏懼。」 國衛院群體健康科學研究所所長 邱弘毅



成功大學善用 NetApp 全快閃設備，駕馭 AI 發展及資料爆量趨勢

評估重點：符合關鍵應用、投資不浪費、不增加管理負擔、服務不中斷

文/ 廠商新聞稿 | 2024-01-19 發表

讚 50

分享



成功大學計算機與網路中心 李南逸組長（左）、陳培殷主任（中）、黃琮富資深網路管理師（右）

iT+ 看影片追技術



如何以 Cloud Native 工具服務打造您的雲端安全

Cloud Summit 臺灣雲端大會
| 29 分



在 DevOps 的路上測試該怎麼進行

DevOpsDays | 42 分



企業資安事件化繁為簡 必備良方 Cortex XDR

IT EXPLAINED | 40 分



【SOAR資安自動化協作】

安基資訊 | 資安防護服務・企業

THANK YOU

